DOCUMENT RESUME

ED 306 235 TM 012 995

AUTHOR Huntley, Renee M.; Welch, Catherine
TITLE The Effect of Answer Location on Item Difficulty and Discrimination in Language-Usage Tests.
PUB DATE 88
NOTE 21p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Entrance Examinations; *Difficulty Level; Distractors (Tests); Evaluation Methods; High School Students; *Language Tests; Language Usage; *Multiple Choice Tests; Secondary Education; Test Construction; *Test Format; Test Items; Text Structure
IDENTIFIERS American College Testing Program; *Item Discrimination (Tests); *Item Position (Tests); Text Processing (Reading)

ABSTRACT

A study compared student performance on language-usage test items embedded in a passage when the location of the answer was varied. American College Testing (ACT) Assessment experimental units were constructed that presented 35 items whose sequence of foils was varied so that each foil appeared once as the "no change" option embedded in the passage. Each of the four experimental units--tests A, B, C, and D--was administered to approximately 300 examinees as part of the ACT tryout procedures. Additional experimental units, arranged so that the two randomly selected items presented the keyed response or the leading distractor in the first position or in the third or fourth position, were administered to approximately 65 high school students representative of the usual ACT examinee. Examinees were attracted to the keyed response in the first, or "no change" position. Results for the low ability students indicate that students tended to use the first two alternatives as wrong answer options more often. However, overall results suggest that item difficulty was not affected to any significant degree by the position of the key, although some language classifications seemed more susceptible than others. Six tables summarize the effects of answer location. Appendices A and B contain sample items from the tests. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED306235

2.2 Educational assessment and evaluation

The Effect of Answer Location on Item Difficulty and Discrimination in Language-Usage Tests

Renee M. Huntley
American College Testing Program

Catherine Welch
American College Testing Program

ABSTRACT

This study compared student performance on language-usage test items embedded in a passage when the location of the answer was varied. ACT Assessment experimental units were constructed that presented thirty-five items whose sequence of foils was varied so that each foil appeared once as the "NO CHANGE" option embedded in the passage. Although results suggest that item difficulty may not generally be affected to any significant degree by the position of the key, some language classifications seem more susceptible than others.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Renee M. Huntley

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM012995

BEST COPY AVAILABLE

The Effect of Answer Location on Item Difficulty and Discrimination in Language-Usage Tests

INTRODUCTION

Although researchers have devoted considerable effort to examining the effects on performance of locating a keyed response among a set of options, results have been contradictory (Hopkins and Hopkins, 1964; Friel and Johnstone, 1979; Fagley, 1987). In general, very little evidence has been uncovered to support the notion that the position of the key affects performance (Cronbach, 1950; Marcus, 1963; Hershey and Hill, 1976). Some researchers have even suggested that it is "unnecessary to follow elaborate schemes to achieve a balanced distribution of keyed responses" since "it would appear that examinees pay less heed to response patterning than might be supposed" (Jessell and Sullins, 1975). Most of the research in this area, however, has examined this phenomenon "in personality, interest, and attitude scales... surprisingly little work has been reported in the case of objective tests of intellectual material" (Gustav, 1963). Applying the research to these tests of educational achievement might raise the same tantalizing questions about keyed-response position because the questions inhere in multiple-choice testing. Or the questions might arise because of the content domain being tested or because of some unique variation in format. Let us consider the problem in the context of language tests.

Multiple-choice language tests can be constructed in varying formats. One format widely used by ACT embeds items within a passage: a continuous piece of discourse having underlined and numbered portions is presented in one column. In a matching column, alternative responses are provided from which the examinee must choose the one alternative that best expresses the idea, makes the statement appropriate for standard written English, or is worded most consistently with the style and tone of the passage as a whole. In this

passage-related format, the language situation being tested in the underlined segment in the passage column is represented as the first among the options in the response column where it is labelled "NO CHANGE." Thus, if the examinee reads through the passage and determines that this first version is indeed the best answer, the examinee would select from the response column the option "NO CHANGE" as the answer and mark the answer sheet accordingly.

The passage is essential to this mode of presentation: items cannot generally be determined without reading the passage. This dependence on the passage, intuition would suggest, might result in special effects associated with this first, "NO CHANGE", answer option -- where the version is embedded in the passage -- such that examinees would be directed to or misdirected from the keyed answer. To the extent that such a phenomenon could be substantiated empirically, then answer location could conceivably affect student performance and content validity. This concern obviously arises in this study from a content perspective and because of a rather unique testing format.

METHOD

Phase 1

Four experimental tryout units -- Tests A, B, C, and D -- were constructed so that each contained thirty-five passage-related test items. Each of the four possible answer options (i.e., the keyed response plus the three distractors) was systematically embedded in the passage (and thus became the "NO CHANGE" option) across these four tryout units. (See Appendix A for examples.) Thus the keyed response for each item was always one variant in

the "NO CHANGE" position, and other arrangements of the same items might vary the location of the key among the second, third, and fourth positions. In this particular phase of the study, however, the key was not systematically varied across each of the second, third, and fourth positions.

Each experimental unit was administered to approximately 300 examinees (in randomly equivalent samples such that only one version was administered to any one examinee) as part of the ACT Assessment tryout procedures.

The data were initially analyzed by using the item analysis results from each of the four versions of each item. The difficulty values (percentage of examinees answering an item correctly) of different versions of an item were studied by using the group as a total as well as by dividing the group into three separate ability levels based on total test performance. The ability levels included a high group (upper 27%), a middle group (46%), and a low group (lower 27%) of examinees. The discrimination values (biserial correlations) resulting from the same item analysis were also studied.

A single-factor design was used to investigate the effect on difficulty of embedding the correct response (keyed response) in the text. A separate analysis was used to investigate the effect on difficulty of embedding the leading distractor in the text. A third analysis was performed to investigate the effect on discrimination of embedding the correct response in the text. A fourth analysis was performed that treated the content classification of an item as a fixed effect, and the effect on difficulty of embedding the correct response in the text was studied. Three broad content classifications were used to identify the items. They were sentence structure; punctuation and grammar; and logic, style, and organization.

The Effect of Answer Location on Item Difficulty and Discrimination in Language-Usage Tests

Phase 2

An additional four experimental tryout units -- Tests 1C, 2D, 3F, and 4G -- were constructed by slightly modifying the original four experimental tryouts. In these new units, four items were further manipulated in each tryout so that two randomly selected items presented the keyed response twice (once in the first position embedded in the passage and once in either the third or fourth position), and two other randomly selected items presented the leading distractor twice (once in the first position embedded in the passage and once in either the third or fourth position). (See Appendix B for examples.)

Each of these experimental units was administered in high school classes to approximately sixty-five examinees in randomly equivalent samples such that only one version was administered to any one examinee. Six local high schools that were reasonably representative of the typical ACT examinee were identified in two midwestern states. In the fall of the school year, the tests were administered to intact classes of seniors who were likely to pursue some form of postsecondary education; the time allotted for the examinees was the same as that allotted for regular ACT administrations. The students were told that the study was designed to determine whether alternative formats in presenting test items would affect students' scores.

Item analysis results for total group performance as well as for the three ability-level groups were also used in this phase to study the effect of repeating the keyed response or the leading distractor twice, once in the first ("NO CHANGE") position and again in the third or fourth position.

Phase 3

In this phase of the study, seven additional tryout units were prepared. One tryout served as the base from which three tryout units were constructed that systematically varied the position of the key across all option positions, and three tryout units were constructed that systematically varied the position of the leading distractor across all option positions.

Each of these experimental units will be administered to approximately 300 examinees (in randomly equivalent samples such that only one version will be administered to any one examinee) as part of the ACT Assessment tryout procedures at some administration later this year.

RESULTS

Phase 1

Total group performance was examined for versions of the items that varied the position of the keys. Table 1 shows the mean proportion of examinees choosing a specific key position across all items. When the key appeared in the second position, the items were less difficult overall than when the key appeared in the first ("NO CHANGE") position. When the key appeared in the third or fourth position, the items became slightly more difficult than when the key appeared in the first ("NO CHANGE") position. However, the proportion of examinees choosing the key was not significantly affected by the position of the key; only nine items were easier for total group performance when the key appeared in the first ("NO CHANGE") position than they were for total group performance when the key appeared in any of the other three positions.

The Effect of Answer Location on Item Difficulty and Discrimination in Language-Usage Tests

Total group performance on versions of the items that varied the position of the leading distractor was also examined. The proportion of the examinees selecting the leading distractor was not affected by the position of that distractor. In only nine of the items where the leading distractor was in the first ("NO CHANGE") position did the leading distractor draw a slightly greater proportion of examinees. Since the items generally had very high difficulty values, it was impossible to identify a pattern among the second and third distractors. Therefore, the second and third distractors were not examined on an item-by-item basis. These results are summarized in Table 2.

To further study the effects on the key and the leading distractor, the total group performance was separated into three ability-level groups based on total test performance. The results were very similar to those for total group performance. For the high- and middle-ability groups, a greater proportion of examinees was drawn to the key in the first ("NO CHANGE") position on ten items than was drawn to the key in any of the other three positions for these items. For the low-ability group, a greater proportion of the examinees was drawn to the key when it appeared in the first ("NO CHANGE") position on eight items than was drawn to the key in any of the other three positions for these items. The leading distractor did not draw a significantly greater proportion of examinees from any ability group.

The discrimination of the items was not affected by the position of the key. Table 3 shows the mean discrimination values when the key was varied among the four positions.

The Effect of Answer Location on Item Difficulty and Discrimination in Language-Usage Tests

The results of the analysis of variance are summarized in Tables 4-7. Tables 4-6 indicate that the position main effects were not significant for either the difficulty or the discrimination of the items. This was true when the key was embedded in the first ("NO CHANGE") position as well as when the leading distractor was embedded in the first ("NO CHANGE") position.

The content classification main effect was significant at the .10 level. These results are reported in Table 7. The content classification by position interaction was not significant.

Phase 2

The results for total group performance of repeating the keyed response twice, once in the first ("NO CHANGE") position and again in the third or fourth position, were significant. On the average for the eight items, 73% of the examinees selected the key in its first ("NO CHANGE") position, and 14% selected the key in its third or fourth position. The results for the three ability-level groups were as follows: 85% of the high-, 73% of the middle- and 64% of the low-ability groups selected the key in its first ("NO CHANGE") position. Eleven percent of the high-, 15% of the middle-, and 17% of the low-ability groups selected the key in its third or fourth position.

The results for total group performance of repeating the leading distractor twice, once in the first ("NO CHANGE") position, and again in the third or fourth position, were not significant. For the eight items, 9% of the examinees selected the leading distractor in the first ("NO CHANGE") position; 5% of the examinees selected the leading distractor in the third or fourth position. The results for the three ability-level groups were as follows: 2% of the high-, 7% of the middle-, and 19% of the low-ability

groups selected the leading distractor in the first ("NO CHANGE") position. One percent of the high-, 5% of the middle-, and 9% of the low-ability groups selected the leading distractor in the third or fourth position.

DISCUSSION

It seems clear from the Phase 2 study, in which the keyed response and leading distractor were presented twice, that examinees were significantly attracted to the keyed response in the first ("NO CHANGE") position. Perhaps this finding could bolster the intuition that answer location in this unique language-test format might affect student performance. As regards the leading distractor, the results for the low-ability group in particular lend credence to Gustav's finding that "students tended to make greater use of... A and B as wrong answer options and to make lesser use of... C and D alternatives as wrong answer options" (Gustav 1963).

Regarding the issue of answer location and content validity, of the items with keyed response in the first ("NO CHANGE") position that were easier for one or more of the three ability-level groups of examinees, item classifications were as follows: seven Sentence Structure items, three Logic and Organization items, one Punctuation item, two Grammar items, and one Diction and Style item. The first-position keyed responses of four of these items were chosen in greater proportions by all three of the groups; the first-position keyed responses of five of these items where chosen in greater proportion by the high or high/middle groups; and the first-position keyed responses of six of these items were chosen in greater proportion by the middle, low, or middle/low groups.

The Sentence Structure items that were thus affected were ones that depended on selecting the response that would complete the parallelism being set up by one piece of a pattern. When the key is in the first position (that is, embedded in the passage), the parallelism is more visible. From the examinee's point of view, it may become increasingly difficult to hold in mind the pieces of the pattern as the examinee checks each subsequent foil. This might explain the finding that placing the key in the second position seemed to make the item easier than did placing it in the third or fourth position.

Even the Logic and Organization items that were affected seemed to depend on the increased visibility of pieces of a pattern when the key was in the first position (embedded in the passage), and one or two sentences before or after the item were part of that pattern.

Logic and Organization and Sentence Structure items that appeared unaffected seemed to be testing the larger, more global concerns of the essay as a whole than did the affected items. In the first case, no paragraphing items were affected; in the latter case, no items dealing with the larger problem of consistency of person throughout the essay were affected.

Perhaps, given the nature of discourse, these results are not entirely unexpected. Although the special effects of locating the answer in the first position (i.e., for the test units in this study, embedding the keyed response within the passage) do not appear significant in the Phase 1 study, except for certain content classifications, it might behoove the test

constructor to consider the results of the Phase 2 study and vary the position of the key in items like those above that did show some effect. The results of the Phase 3 study, which will be examined once the data are collected, may shed further light on the matter, as may a planned investigation of gender differences across the three phases.

References

Cronbach, L. J. (1950). Further evidence on response sets and test design. Educational and Psychological Measurement, 10, 3-31.

Fagley, N. S. (1987). Positional response bias in multiple-choice tests of learning: its relation to testwiseness and guessing strategy. Journal of Education Psychology, 79, 95-97.

Friel, S., & Johnstone, A. H. (1979). Does the position matter? Education in Chemistry, 16, 175.

Gustav, A. (1963). Response set in objective achievement tests. Journal of Psychology, 56, 421-427.

Hershey, M. R., & Hill, D. B. (1976). Positional response set in pre-adult socialization surveys. Social Science Quarterly, 56, 707-714.

Hopkins K. S., & Hopkins, B. R. (1964). Intraindividual and interindividual positional preference response styles in ability tests. Educational and Psychological Measurement, 24, 801-805.

Jessell, J. C., & Sullins, W. L. (1975). The effect of keyed response sequencing of multiple choice items on performance and reliability. Journal of Educational Measurement, 12, 45-48.

Marcus, A. (1963). The effect of correct response location on the difficulty level of multiple-choice questions. Journal of Applied Psychology, 47, 48-51.

The Effect of Answer Location on Item Difficulty and Discrimination in Language-Usage Tests

Table 1

Proportion of Examinees Selecting the Key when Varied Among Positions

	"NO CHANGE" 1st position	2nd position	3rd position	4th position
$\bar{X}$	72.35	74.51	68.26	69.82
SD	14.97	9.97	17.47	16.19

Table 2

Average Proportion of Examinees Selecting the Leading Distractor When Varied Among Positions

	"NO CHANGE" 1st position	2nd position	3rd position	4th position
$\bar{X}$	16.20	17.94	16.20	19.16
SD	9.90	9.10	8.80	12.57

Table 3

Average Discrimination Value of Items When Key is Varied Among Positions

	"NO CHANGE" 1st position	2nd position	3rd position	4th position
$\bar{X}$	47.50	47.64	45.25	44.87
SD	10.98	15.06	12.76	9.87

The Effect of Answer Location on Item Difficulty and Discrimination in Language-Usage Tests

Table 4

ANOVA Summary Table:
Analysis of Key-Position Effect on Difficulty

Source	SS	df	MS	F
Position	229	3	76	.38
Error	26,520	132	201	
Total	26,749	135		

Table 5

ANOVA Summary Table:
Analysis of Leading-Distractor-Position Effect on Difficulty

Source	SS	df	MS	F
Position	77	3	26	.24
Error	14,540	136	107	
Total	14,617	139		

Table 6

ANOVA Summary Table:
Analysis of Key-Position Effect on Discrimination

Source	SS	df	MS	F
Position	206	3	69	.45
Error	20780	137	152	
Total	20986	140		

The Effect of Answer Location on Item Difficulty and Discrimination in Language-Usage Tests

Table 7

ANOVA Summary Table:
Analysis of Position by Content Classification

Source	SS	df	MS	F	F prob.
Content Classification	872	2	436	2.31	.10
Position	233	3	78	.41	
Content Classification x Position	1,054	6	176		
Error	22,722	120	189		
Total	24,881	131			

The Effect of Answer Location on Item Difficulty and Discrimination in Language-Usage Tests

APPENDIX A

Phase 1 Examples

Version A

Passage I

Unlike many wild species in danger of extinction, the coyote is flourishing. Once confined to the West, coyotes can now be found throughout the United States. They raised[1] garbage cans in New York suburbs and drinking its swimming pools.[2] Why has the coyote survived while such more larger, stronger[3] relatives as the timber wolf have nearly disappeared? The answer lies in the coyote's unique adaptability. Whereas most creatures have rigid feeding habits. The coyote is virtually omnivorous.[4] In wilderness areas, the coyote pursues a variety of prey: prairie dogs, rabbits, young deer, or even fierce badgers.[5] Nearer to civilization, it dines on kitchen scraps or an occasional rodent.

Coyotes are resourceful not only in finding food and[6]

1. A. NO CHANGE
   B. have raided
   C. have been seen raiding
   D. have been seen to have raided

2. F. NO CHANGE
   G. its beautifully tiled swimming pools.
   H. from swimming pools with beautiful tile in Hollywood.
   J. from swimming pools in Hollywood.

3. A. NO CHANGE
   B. larger, stronger
   C. larger stronger
   D. more larger, more stronger

4. F. NO CHANGE
   G. habits. Nevertheless, the
   H. habits, the
   J. habits, but the

5. A. NO CHANGE
   B. even fierce badgers are pursued.
   C. fierce badgers are pursued.
   D. pursues even fierce badgers.

6. F. NO CHANGE
   G. in finding food but also
   H. either in finding food and
   J. not only in finding food but also

The Effect of Answer Location on Item Difficulty and Discrimination in Language-Usage Tests

APPENDIX A (cont.)

Version B

Passage I

Unlike many wild species in danger of extinction, the coyote is flourishing. Once confined to the West, coyotes can now be found throughout the United States. They have raided[1] garbage cans in New York suburbs and drinking its beautifully tiled swimming pools.[2]

Why has the coyote survived while such larger, stronger[3] relatives as the timber wolf have nearly disappeared? The answer lies in the coyote's unique adaptability. Whereas most creatures have rigid feeding habits. Nevertheless,[4] the coyote is virtually omnivorous. In wilderness areas, the coyote pursues a variety of prey: prairie dogs, rabbits, young deer, or even fierce badgers are pursued.[5] Nearer to civilization, it dines on kitchen scraps or an occasional rodent.

Coyotes are resourceful in finding food but also[6] in capturing it. They often hunt in pairs.[7] As one chasing the prey while the other waits in ambush.

Most important, however, is the coyotes[8] keen sense

1. A. NO CHANGE
   B. raided
   C. have been seen raiding
   D. have been seen to have raided

2. F. NO CHANGE
   G. its swimming pools.
   H. from swimming pools with beautiful tile in Hollywood.
   J. from swimming pools in Hollywood.

3. A. NO CHANGE
   B. more larger, stronger
   C. larger stronger
   D. more larger, more stronger

4. F. NO CHANGE
   G. habits. The
   H. habits, the
   J. habits, but the

5. A. NO CHANGE
   B. even fierce badgers.
   C. fierce badgers are pursued.
   D. pursues even fierce badgers.

6. F. NO CHANGE
   G. not only in finding food and
   H. either in finding food and
   J. not only in finding food but also

7. A. NO CHANGE
   B. pairs
   C. pairs, as
   D. pairs, with

8. F. NO CHANGE
   G. their
   H. its
   J. it's

The Effect of Answer Location on Item Difficulty and Discrimination in Language-Usage Tests

APPENDIX A (cont.)

Version C

Passage I

Unlike many wild species in danger of extinction, the coyote is flourishing. Once confined to the West, coyotes can now be found throughout the United States. They have been seen raiding[1] garbage cans in New York suburbs and drinking from swimming pools with beautiful tile in Hollywood.[2]

Why has the coyote survived while such larger stronger[3] relatives as the timber wolf have nearly disappeared? The answer lies in the coyote's unique adaptability. Whereas most creatures have rigid feeding habits,[4] the coyote is virtually omnivorous. In wilderness areas, the coyote pursues a variety of prey: prairie dogs, rabbits, young deer, or fierce badgers are pursued.[5] Nearer to civilization, it dines on kitchen scraps or an occasional rodent.

Coyotes are resourceful either in finding food and[6] in capturing it. They often hunt in pairs,[7] as one chasing the prey while the other waits in ambush. Most important, however, is its[8] keen sense

1. A. NO CHANGE
   B. have raided
   C. raided
   D. have been seen to have raided

2. F. NO CHANGE
   G. its beautifully tiled swimming pools.
   H. its swimming pools.
   J. from swimming pools in Hollywood.

3. A. NO CHANGE
   B. larger, stronger
   C. more larger, stronger
   D. more larger, more stronger

4. F. NO CHANGE
   G. habits. Nevertheless, the
   H. habits. The
   J. habits, but the

5. A. NO CHANGE
   B. even fierce badgers are pursued.
   C. even fierce badgers.
   D. pursues even fierce badgers.

6. F. NO CHANGE
   G. in finding food but also
   H. not only in finding food and
   J. not only in finding food but also

7. A. NO CHANGE
   B. pairs. As
   C. pairs
   D. pairs, with

8. F. NO CHANGE
   G. the coyotes
   H. their
   J. it's

The Effect of Answer Location on Item Difficulty and Discrimination in Language-Usage Tests

APPENDIX A (cont.)

Version D

Passage I

Unlike many wild species in danger of extinction, the coyote is flourishing. Once confined to the West, coyotes can now be found throughout the United States. They have been seen to have raided[1] garbage cans in New York suburbs and drinking from swimming pools in Hollywood.[2]

Why has the coyote survived while such more larger, more stronger[3] relatives as the timber wolf have nearly disappeared? The answer lies in the coyote's unique adaptability. Whereas most creatures have rigid feeding habits, but the[4] coyote is virtually omnivorous. In wilderness areas, the coyote pursues a variety of prey: prairie dogs, rabbits, young deer, or pursues even fierce badgers.[5] Nearer to civilization, it dines on kitchen scraps or an occasional rodent.

Coyotes are resourceful not only in finding food but also[6] in capturing it. They often hunt in pairs, with[7] one chasing the prey while the other waits in ambush. Most important,

1. A. NO CHANGE
   B. have raided
   C. have been seen raiding
   D. raided

2. F. NO CHANGE
   G. its beautifully tiled swimming pools.
   H. from swimming pools with beautiful tile in Hollywood.
   J. its swimming pools.

3. A. NO CHANGE
   B. larger, stronger
   C. larger stronger
   D. more larger, stronger

4. F. NO CHANGE
   G. habits. Nevertheless, the
   H. habits, the
   J. habits. The

5. A. NO CHANGE
   B. even fierce badgers are pursued.
   C. fierce badgers are pursued.
   D. even fierce badgers.

6. F. NO CHANGE
   G. in finding food but also
   H. either in finding food and
   J. not only in finding food and

7. A. NO CHANGE
   B. pairs. As
   C. pairs, as
   D. pairs

The Effect of Answer Location on Item Difficulty and Discrimination in Language-Usage Tests

APPENDIX B

Phase 2 Examples

Version 1C (Repeated Key)

Why has the coyote survived while such

larger, stronger relatives as the timber
3

wolf have nearly disappeared? The answer

lies in the coyote's unique adaptability.

3. A. NO CHANGE
   B. larger stronger
   C. larger, stronger
   D. more larger, more stronger

Version 2D (Repeated Leading Distractor)

caught. Thus, despite attempts to eliminate coyotes

by ranchers similar to city-dwellers, their remarkable
10

adaptability has almost certainly ensured their survival.

10. F. NO CHANGE
    G. likewise to city-dwellers,
    H. similar to city-dwellers,
    J. and city-dwellers alike.